Task-Independent Workload Classification using Non-Binary Output and Task Scaling

Laurens R. Krol[1,4], Maya-Christin Müller[2], Antoniya A. Boyanova[3], Andrea Kübler[2], Thorsten O. Zander[1,4]

1) Neuroadaptive Human-Computer Interaction, BTU Cottbus-Senftenberg, Germany
2) Department of Psychology, Julius-Maximilians-Universität of Würzburg, Germany
3) School of Psychology, University of Aberdeen, Scotland, UK
4) Zander Labs GmbH, Germany

Keywords: Workload, Classification, Task-Independence

Introduction

Workload remains one of the most sought-after EEG-based measures in the context of neuroergonomics and neuroadaptive technology, given the high potential utility of its real-time detection in productive environments [1]. For real-world applications, a workload classifier that does not require (re-)calibration across different contexts, tasks, or users is of particular interest (e.g., [2]). In previous work, we have developed a workload calibration paradigm that calibrates in under ten minutes, with a corresponding classifier that appears to function well across tasks [3,4], particularly when looking at continuous rather than binary classifier output values [5]. Here, we expand on this work with a new selection of tasks and additional "scaling tasks" to further improve performance.

As shown in [5] and [6], a continuous interpretation of classifier output can help when translating values between tasks or contexts. For example, imagine a classifier that is calibrated on "high" versus "low" workload, where "high" load is induced using a particularly difficult arithmetic task. Because of this difficulty, both "high" and "low" conditions in different, subsequent task may both be relatively easy in comparison, resulting in a "low" workload detection by the classifier regardless of the actual condition. When looking at continuous classifier outputs rather than binary categorizations, however, this same classifier may still produce statistically significant differences between the two conditions. In other words, when the threshold between "high" and "low" is 0.5, interpreting both .1 and .4 simply as "low" ignores the meaningful difference that exists between them.

The effective workload induced by a standard calibration task can differ per participant depending on their skill or experience with respect to that task. The range of mental exertion thus captured by a classifier will differ, and may not correspond to or even include the range of exertion that that same individual would experience during a different task. To account for this, we propose using additional tasks to quantify participants' skills. For this experiment, we hypothesized that different tasks induce workload to different extents, that a continuous interpretation of classifier output reflects these differences better than a binary interpretation, and that the aforementioned quantifications can additionally scale the classifier output to cover a range appropriate for different tasks and skills.

Participants, Tasks, and Methods

20 participants aged 20-39 were first given four scaling tasks to measure their general ability in four areas: mental arithmetic using a version of the computer-based MATH test [7], spatial cognition

using a pen-and-paper mental rotation test [8], linguistic ability using the German, pen-and-paper *Mehrfachwahl-Wortschatz Test B* [9], and short-term memory span using a computer-based version of the corresponding part of the Wechsler Adult Intelligence Scale IV [10]. Following this, 64-channel EEG (Brain Products actiCHamp) was recorded while participants performed the calibration task previously described in [3] and [5]. Then, in random order, they performed five more tasks: 1. An addition task (AD) requiring them to add two three-digit numbers in either high or low difficulty conditions (Q-values [11] >4 or <2); 2. A word recovery task (WR) where participants were shown words with letters in random order and were required to "unscramble" them, these words being longer, less common words (high workload) or shorter, more common words (low workload); 3. A mental rotation task (MR) using stimuli from [12], with difficulty varied by rotating the stimuli to different degrees across two planes (high) or one plane (low); 4. A backward digit span task (BDS) where participants were given a sequence of 5 (high) or 3 (low) digits one after the other and asked to reproduce it in reverse order; 5. An n-back task (NB) [13] with n=2 (high) and n=1 (low).

We trained individual classifiers for each participant on the data from the calibration task, and applied it to their data from the remaining five tasks. For this, data from each of the variable-length trials was segmented into 1-second epochs resulting in at least 150 epochs per class per task. A filter-bank common spatial patterns (FB-CSP, [14]) approach was used using frequency bands 4-7 and 8-13 Hz. The top three filter pairs were used to train a classifier using regularized linear discriminant analysis (LDA). In a first analysis, the trained classifier was applied in a binary fashion to the data of the five tasks. In a second analysis, the raw classifier output was used instead and inserted as dependent variable into a 2x5 (condition x task) repeated-measures analysis of variance (rmANOVA) across all participants. Bonferroni-adjusted pairwise comparisons were calculated to test for effects.

Results and Conclusion

For the first, binary analysis, average accuracies across participants for the five tasks were not significant, ranging between 46 and 50%. This was expected, as explained above. The rmANOVA results, however, showed a significant main effect of condition ($F(1, 19) = 25.59$, $p < .001$, $\eta_p^2 = .57$), indicating that "high" conditions did result in significantly higher classifier output across all five tasks.

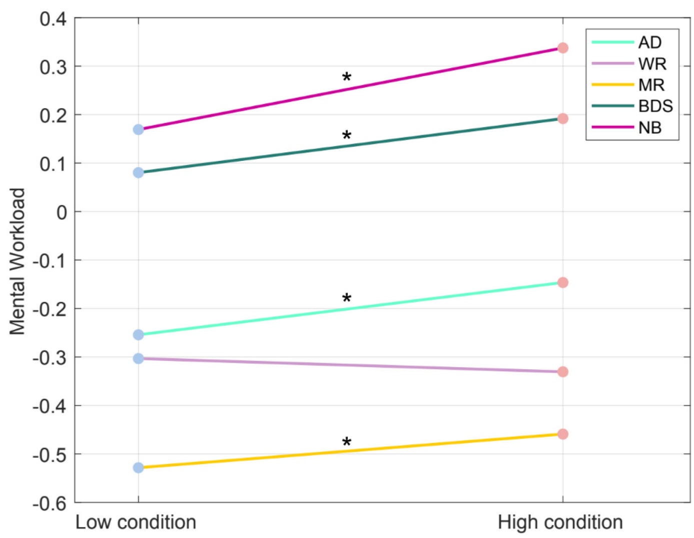

Fig. 1: Simple effects of factor condition *from the rmANOVA. Asterisks indicate significance at, at least, α = 0.05.*

Fig. 1 illustrates the different values for the five tasks separately. Some tasks, e.g. the n-back, appear to consistently induce a higher load than e.g. the mental rotation task, which would confound binary classification, but still show significant differences between conditions. Pairwise comparison tests indicated that this effect was significant for all tasks but WR.

At the time of writing, the scaling tasks have not yet been included in the analysis, but the different levels of overall average workload across tasks, as seen in the figure, indicate that a scaling is useful. The next step is thus to evaluate how the participants' scores on the scaling tasks correlate with their workload levels across tasks.

All in all, the results provide additional indications that a continuous approach, rather than a binary ("low" versus "high") approach to workload classification is a viable method to obtain meaningful workload measures from EEG.
(This work was supported by the U.S. Air Force, FA8655-20-1-7007.)

References

1. Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58*(1), 1–17. doi: 10.1080/00140139.2014.956151
2. Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience, 8*, 114. doi: 10.3389/fnins.2014.00114
3. Krol, L. R., Freytag, S.-C., Fleck, M., Gramann, K., & Zander, T. O. (2016). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 003171–003174). doi: 10.1109/SMC.2016.7844722
4. Andreessen, L. M., Gerjets, P., Meurers, D., & Zander, T. O. (2021). Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. *User Modeling and User-Adapted Interaction, 13,* 75–104. doi: 10.1007/s11257-020-09273-5
5. Zhang, X., Krol, L. R., & Zander, T. O. (2018). Towards task-independent workload classification: Shifting from binary to continuous classification. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 556–561). doi: 10.1109/SMC.2018.00104
6. Krol, L. R., Klaproth, O. W., Vernaleken, C., Russwinkel, N., & Zander, T. O. (2022). Towards neuroadaptive modeling: assessing the cognitive states of pilots through passive brain-computer interfacing. In S. H. Fairclough & T. O. Zander (Eds.), *Current research in neuroadaptive technology* (pp. 59–73). London, UK: Academic Press. doi: 10.1016/B978-0-12-821413-8.00009-9
7. Turner, J. R., Hewitt, J. K., Morgan, R. K., Sims, J., Carroll, D., & Kelly, K. A. (1986). Graded mental arithmetic as an active psychological challenge. *International Journal of Psychophysiology, 3*(4), 307–309. doi: 10.1016/0167-8760(86)90039-5
8. Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test - different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58. doi: 10.1006/brcg.1995.1032
9. Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B* (5. Auflage). Spitta GmbH.
10. Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV)*. APA PsycTests. doi: 10.1037/t15169-000
11. Thomas, H. B. G. (1963). Communication theory and the constellation hypothesis of calculation. *Quarterly Journal of Experimental Psychology, 15*(3), 173–191. doi: 10.1080/17470216308416323
12. Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition, 66*(3), 260–264. doi: 10.1016/j.bandc.2007.09.003
13. Kirchner, W. L. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358. doi: 10.1037/h0043688
14. Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *IEEE International Joint Conference on Neural Networks (IJCNN), 2008* (pp. 2390–2397). doi: 10.1109/IJCNN.2008.4634130

The Third Neuroadaptive Technology Conference

Conference Programme